稀疏和标签约束半监督自动编码机的分类算法 *

王慧玲 [1]，宋 威 [1,2]，王晨妮

(1. 江南大学 物联网工程学院，江苏 无锡 214122; 2. 物联网技术应用教育部工程研究中心，江苏 无锡 214122)

摘 要：自动编码机通过深度无监督学习能够表达数据的语义特征，但由于其隐含层节点个数难以有效确定，所处理的数据进一步用于分类时常会导致分类准确度低，稳定性弱等问题。针对这些问题，提出了一种稀疏和标签约束的半监督自动编码机(SLRAE)，以实现无监督学习与监督学习的有机结合，更准确地抽取样本的本质特征。稀疏约束项针对每个隐含结点的响应添加约束条件，从而在隐含神经元数量较多的情况下仍可发现数据中潜在的结构；同时引入标签约束项，以监督学习的方式比对实际标签与期望标签，针对性地调整网络参数，进一步提高分类准确率。为验证所提方法的有效性，实验中对多个数据集进行广泛地测试，其结果表明，相对传统自动编码机(AE)，稀疏自动机（SAE)，以及极限学习机(ELM)，SLRAE 所处理的数据应用于同一分类器，能明显提高分类准确率和稳定性。

关键词：分类；稀疏约束；标签约束；自动编码机；极限学习机

中图分类号：TP301.6　　doi: 10.3969/j.issn.1001-3695.2018.02.0147

Semi-supervised auto-encoder using sparse and label regularizations for classification

Wang Huiling[1], Song Wei[1, 2], Wang Chenni[1, 2]

(1. *School of Internet of Things Engineering*, *Jiangnan University*, *Wuxi Jiangsu* 214122, *China*; 2. *Engineering Research Center of Internet of Things Technology Applications for Ministry of Education*, *Wuxi Jiangsu* 214122, *China*)

Abstract: Auto-Encoder can express the semantic features of data through deep unsupervised learning, but it is hard to determine the nodes of hidden layer and the processing of data for classification often leads to low accuracy and low stability. To solve the problems, this paper proposes a semi-supervised auto-encoder using sparse and label regularizations (LSRAE) to extract the essential features of the samples more accurately by combining unsupervised learning with supervised learning. The sparse regularization term adds constraints to the response of each hidden node, so that this algorithm can find potential structures in the data when the number of hidden neurons is large. At the same time, this algorithm introduce a label regularization term to compare the actual labels with desired labels by supervised learning to adjust the network parameters and further improve the classification accuracy. In order to verify the validity of the proposed method, this algorithm tests many data sets in the experiment. The results show that compared with traditional auto-encoders (AE) , sparse auto-encoder (SAE) , and extreme learning machine (ELM) , SLRAE can obviously improve the classification accuracy and stability when the processed data is applied to the same classifier.

Key words: classification; sparse regularization; label regularization; auto-encoder; extreme learning machine

0 引言

随着信息技术的飞速发展，数字数据库中的数据呈现指数增长，如何有效地从海量的信息中提取出有价值的特征，已成为当今数据挖掘和模式识别领域的研究热点之一。自 2006 年以来，以特征学习[1]为代表的自动编码器（auto-encoder，AE）[2]表示学习在机器学习领域取得了突破性的进展，这些进展主要用来进行深度学习模型的初始化[3]。然而，常用的 AE 会因其复杂的网络结构而导致分类精度下降。为了解决这一问题，一些 AE 正则化方法被学者们提出：稀疏自动编码器（sparse auto-encoder，SAE）[4,5]是 Bengio 等人于 2007 年提出来的，要求隐藏层神经元的激活度满足一定的稀疏性；降噪自动编码器（denoising auto-encoder，DAE）[6,7]是由 Vincent 等人于 2008 年提出的，是对输入向量加入干扰，训练编码器重构出原始输入，使网络具有更好的鲁棒性；压缩自动编码机（contractive autoencoder，CAE)[8,9]是 Bengio 等人在 2011 年提出，通过将隐

收稿日期：2018-02-26；**修回日期：**2018-04-17　**基金项目：**国家自然科学基金资助项目（61673193）；中央高校本科研业务费专项资金资助项目（JUSRP51635B，JUSRP51510）；江苏省自然科学基金资助项目（BK20150159）；中国博士后科学基金资助项目（2017M621625）

作者简介：王慧玲（1993-），女，江苏盐城人，硕士，主要研究方向为深度学习、人工智能（1159439202@qq.com）；宋威（1981-），男，湖北恩施人，副教授，博士，主要研究方向为数据挖掘；王晨妮（1997-），女，浙江杭州人，本科生，主要研究方向为深度学习、群体智能.

含层输出关于权重的雅克比矩阵的 F 范数的平方作为重构函数的正则化约束项，可以获取鲁棒性的中间层特征。

以上的正则化 AE 都是无监督学习的过程[10]，无监督的深度学习方法是从数据本身分析，能有效提取出深层抽象的数据特征，但是由于先验标签信息的缺乏，所提取的特征不能描述特定类别，很难应用于分类[11]。监督学习就可以很好地解决这个问题，但学习到的特征不能很好地表示原始的数据，泛化能力弱，很容易产生过拟合问题。

为了更好地解决无监督学习和监督学习方法存在的问题，本文提出了一种稀疏和标签约束的半监督自动编码机（SLRAE），用无监督学习中得到的基本语义特征和监督学习中描述标签的有鉴别性的特征来表示数据的特点，再进一步将这些特征用于分类。为了验证我们提出的方法的有效性，我们在 USPS 数据库[12]和 6 个 UCI 数据集[13]上做大量的实验。我们将本文提出的 SLRAE 方法与 AE，SAE，ELM，LRAE(只使用标签约束项的自动编码机)，DBN[14]和 Adaptive DBN[15]做比较，实验结果表明 SLRAE 具有更好的分类准确性。

1 相关工作

1.1 自动编码机(AE)

自动编码机的概念是 20 世纪 80 年代晚期被提出来的，基本的自动编码机由三层神经网络组成，包括输入层、隐含层和输出层，输入层与输出层的神经元的节点数相同。在编码阶段，用函数 f_θ 对输入向量 x 进行映射，获得中间层表示 y，在解码阶段用函数 g_θ 对中间层表示 y 进行映射获得重构数据 z。自动编码机通过最小化重构误差 $L(x,z)=\sum_{i=1}^{n}\|x_i-z_i\|^2$ 来微调网络参数，其中 x_i 是第 i 维训练样本，z_i 是第 i 维输出数据，n 是训练样本的数量。

在 AE 中需要调整的参数是 $\theta=\{W,b,W',b'\}$，其中 b 和 b' 是编码和解码的偏置，W 和 W' 是编码和解码的权重，W' 是 W 的转置。通常采用经典的梯度下降法获得最优值作为参数 θ。与有着较大的人工数据特征提取工作量的传统的 BP 神经网络[16]相比，AE 可以提高特征提取的工作效率，减少原始输入数据的维数，还可以很好地学习到给定数据集的压缩和分布式特征表示。

1.2 极限学习机(ELM)

极限学习机[17,18]是由黄广斌教授提出的一种具有快速学习特点的单隐层前馈神经网络，隐含层输入权值和偏置都是随机产生的，只需要设置隐含层神经元的个数，计算该网络的输出权值。传统的 ELM 包括输入层、隐含层和输出层，假设有 N 个任意的样本 $\{(x_j,t_j)\}_{j=1}^{N}$，该 ELM 网络结构在输入为 x 的隐含层输出矩阵为 H，则整个网络的输出层可表示为

$$H\beta=T \tag{1}$$

其中：β 是隐含层和输出层的权重矩阵 $T=[t_1,t_2,\dots,t_N]_{m\times N}$ 表示样本的目标标签值，这样前馈神经网络的训练就可以转化成一个求解输出权值矩阵的最小二乘解 $\hat{\beta}$ 的问题，输出权值 β 可由下列公式得出：

$$\hat{\beta}=H^{\dagger}T \tag{2}$$

其中 $H^{\dagger}$ 表示隐含层输出矩阵 H 的 Moore-penrose(MP)广义逆。与传统的需要多次迭代反馈的前馈神经网络不同的是，ELM 利用最小二乘法直接计算出权值矩阵，训练速度快，泛化性能好。

2 稀疏和标签约束的半监督自动编码机(SLRAE)

本文提出了一种使用稀疏和标签约束的半监督自动编码机（semi-supervised AE），其中无标签样本的无监督学习过程可以提高半监督自动编码机学习的泛化能力，有标签样本的监督学习过程可以提高半监督自动编码机模型的分类准确度。

在自动编码机中加入标签项，减少训练数据的实际输出标签和期望输出标签之间的误差，从而弥补无监督学习的分类准确度低的缺点。我们还利用稀疏正则化来实现输入数据的稀疏编码，通过约束隐含层节点数从而有效地提取数据的隐藏结构，具有良好的学习数据集的能力，然后将学习到的特征表达输入到分类器中，分类结构验证了所提出的半监督自动编码机的有效性。图 1 显示的是 SLRAE 的用于分类的框架。SLRAE 的代价函数由 J_{AE}，J_{wd}，J_{sparse}，J_{label} 组成。J_{AE} 是输入与输出数据的重构误差，通过最小化 J_{AE} 使得输入数据尽可能地接近输出数据，从而更准确地重构出输出数据。J_{wd} 是为了减小权重的幅度，防止过度拟合。在此基础上，我们所提出的 SLRAE 还包括稀疏约束项 J_{sparse} 和标签误差项 J_{label}。J_{sparse} 限制隐含层使得表达更稀疏，可以使得 SLRAE 有效地从大量的数据中提取出潜在的本质特征。J_{label} 是实际标签与期望标签的误差项，通过使误差项最小调整网络参数，然后将网络输入到分类器中，可以进一步提高分类的准确性。因此，J_{AE}，J_{wd} 和 J_{sparse} 是通过无监督学习获得，J_{label} 是通过有监督学习获得。本文提出的半监督自动编码机 SLRAE 结合了非监督和监督学习的优点，并利用上述 4 个约束来调整权值和偏置，提取出数据的特征表达，并结合分类器，通过构建具有高准确率与高泛化能力的半监督学习网络，进一步实现分类应用。

事实上，我们的标签约束利用监督学习提高分类的准确率，这个标签项是用极限学习机（Extreme Learning Machine，ELM）的思想计算的。由于原始输入数据往往庞大且冗余，难以直接获取本质语义特征，稀疏约束限制隐含层节点数使得表达更稀疏，加入稀疏限制可以获取输入的主要信息，所以本文提出的 SLRAE 不但可以有效地从大量数据中提取潜在的本质特征，还可以提高分类的准确度。

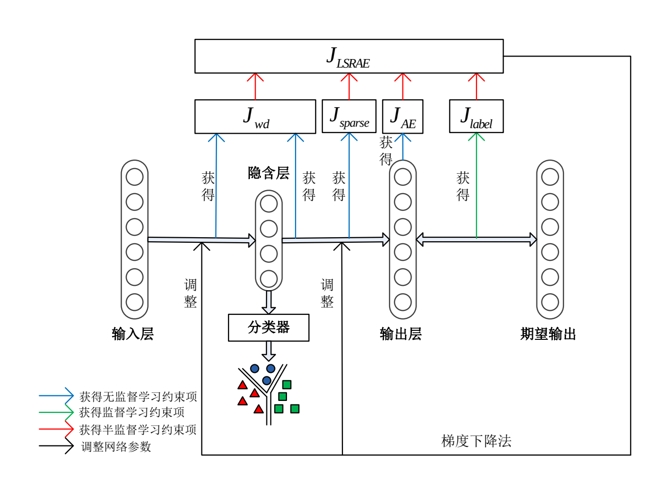


图 1　SLRAE 用于分类的框架

在 SLRAE 中，假设 n 个样本 $X=\{x_i\}_{i=1}^n$ 输入到 d 维空间，每一个样本表示为 $x_i=[x_{i1},x_{i2},...,x_{id}]^T\in R^d$，编码函数 $y_i=f_\theta(x_i)=s(Wx_i+b)$ 将 x_i 映射 l 到维隐含层上得到中间表示 y_i，解码函数 $z_i=g_\theta(y_i)=s(W'y_i+b')$ 重构 x_i 得到 z_i。它们可以简洁地表示为 $a_1=Wx_i+b$，$y_i=s(a_1)$，$a_2=W'y_i+b'$，其中 $s(x)=1/(1+e^{-x})$ 是 sigmoid 激活函数。W 和 b 分别表示输入层与隐含层之间的权重和偏置，W' 和 b' 分别表示隐含层和输出层之间的权重和偏置。我们用 $X\in R^{d\times n}$ 表示输入层的样本特征矩阵，$Y\in R^{l\times n}$ 表示隐含层的样本特征矩阵，$Z\in R^{d\times n}$ 表示输出层对于输入数据的样本特征矩阵。

在本文中，隐含层的节点数远远大于输入层的节点数，我们可以在隐含层上加入一个稀疏限制，它可以保证学习到更多的局部结构信息。在每次迭代，SLRAE 计算出平均激活度 $\rho_j^*=(1/n)\sum_{i=1}^n a_j(x_i)$，其中 n 是样本数，i 是第 i 个样本，$i\in(1,2,3...n)$，$a_j(x_i)$ 表示第 i 个样本情况下，第 j 个隐含神经元的激活度。在稀疏约束条件下，ρ 是稀疏性参数，且 $\rho_j^*=\rho$，通常情况下，ρ 是一个接近于 0 的较小的值。这里，相对熵（KL divergence）是一种用来测量两个分布之间差异的方法，计算公式如下：

$$J_{sparse}=KL(\rho\|\rho_j^*)=\rho\log\frac{\rho}{\rho_j^*}+(1-\rho)\log\frac{1-\rho}{1-\rho_j^*} \quad (3)$$

在本文中也使用了数据的标签，$\{X_n,T_n\}=\{x_i,t_i\}_{i=1}^n$，$t_i=[t_{i1},t_{i2},...,t_{in}]$。$g_i(x)=s(Wx_i+b)$ 是隐含层第 i 个节点的输出值。用 $h(x)=[g_1(x),g_2(x),...g_h(x)]$ 表示输入数据在隐含层的输出，用 H 表示隐含层的输出矩阵 $H=[h^T(x_1),h^T(x_2),...,h^T(x_n)]^T$。因为将自动编码机用于分类前的特征提取，ELM 可以直接用作分类器[19]，所以虚拟出一个标签层，也就是分类的结果，之后将 AE 的输入层和隐含层以及虚拟标签层看做 ELM，然后用 $\hat{\beta}=H^\dagger T$ 计算出隐含层和虚拟标签层之间的权值和偏置。在该 ELM 可以计算出实际的输出标签，公式如下：

$$L=H\hat{\beta} \quad (4)$$

在已知所给的标签 T 和 ELM 计算出的实际输出标签 L 的基础上，得到实际标签与给定标签之间的标签误差，并最小化标签误差 $P(L,T)$，可表示为

$$\min_\theta J_{label}(\theta)=\min_\theta\sum_{i=1}^n P(L,T) \quad (5)$$

那么 SLRAE 的目标函数如下：

$$J_{LSRAE}=J_{AE}+\lambda J_{wd}+\beta J_{sparse}+\gamma J_{label} \quad (6)$$

其中：λ，β 和 γ 是控制公式中各项相对重要性的参数，第一项 $J_{AE}=\frac{1}{2n}\sum_{i=1}^n(\|h_{w,b}(x^{(i)})-y^{(i)}\|^2)$ 是训练样本输入与输出的重构误差项，第二项 $J_{wd}=\frac{1}{2}\sum_{i=1}^d\sum_{j=1}^l(W_{ji})^2$ 是为了防止权重过度拟合的权重衰减项，第三项 $J_{sparse}=\sum_{j=1}^l KL(\rho\|\rho_j^*)$ 是稀疏惩罚项，在隐藏神经元数量较多的情况下仍然可以发现输入数据中重要的结构，最后一项是标签误差项，用来表示原始标签与实际标签的误差，表示如下：

$$J_{label}=\sum_{i=1}^n P(L,T)=\frac{1}{2n}\sum_{i=1}^n\|L-T\|^2=\frac{1}{2n}\sum_{i=1}^n\left\|H\hat{\beta}-T\right\|^2 \quad (7)$$

其中：T 为输入样本的原始标签，H 表示隐含层的输出矩阵，$\hat{\beta}$ 是隐含层与输出标签层的权值与偏置，通过公式 $\hat{\beta}=H^\dagger T$ 计算可得。

用梯度下降法来优化目标函数，更新 W 和 b 公式如下：

$$W_{ij}^{(l)}=W_{ij}^{(l)}-\alpha\frac{\partial J_{LSRAE}(W,b)}{\partial W_{ij}^{(l)}} \quad (8)$$

$$b_i^{(l)}=b_i^{(l)}-\alpha\frac{\partial J_{LSRAE}(W,b)}{\partial b_i^{(l)}} \quad (9)$$

其中：α 为学习率，对于重构误差项 J_{AE}，关于 Z 的偏导数可以表示为 $\frac{\partial J_{AE}}{\partial Z}=-\frac{1}{n}(X-Z)$，那么得到 J_{AE} 关于 b' 和 W' 的偏导数为

$$\frac{\partial J_{AE}}{\partial b'}=\left[\frac{\partial J_{AE}}{\partial Z}\circ s'\left(W'Y+b'\otimes 1_n^T\right)\right]1_n \quad (10)$$

$$\frac{\partial J_{AE}}{\partial W'}=\left[\frac{\partial J_{AE}}{\partial Z}\circ s'\left(W'Y+b'\otimes 1_n^T\right)\right]Y^T \quad (11)$$

相同地，J_{AE} 关于隐含层表示 Y 的偏导数可以表示为 $\frac{\partial J_{AE}}{\partial Y}=W\left[\frac{\partial J_{AE}}{\partial Z}\circ s\left(W'Y+b'\otimes 1_n^T\right)\right]$，那么得到 J_{AE} 关于 b 和 W 的偏导数为

$$\frac{\partial J_{AE}}{\partial b}=\left[\frac{\partial J_{AE}}{\partial Y}\circ s'\left(WX+b\otimes 1_n^T\right)\right]1_n \quad (12)$$

$$\frac{\partial J_{AE}}{\partial W}=\left[\frac{\partial J_{AE}}{\partial Y}\circ s'\left(WX+b\otimes 1_n^T\right)\right]X^T \quad (13)$$

其中：$\circ$ 是阿达玛积，$\otimes$ 是外积，1_n^T 表示数值为 1 的维列向量，

$s'(x)=s(x)[1-s(x)]$ 是 sigmoid 函数的偏导数，我们可以结合(11)和(13)可得

$$\frac{\partial J_{AE}}{\partial W}:=\frac{1}{2}\left\{\left[\frac{\partial J_{AE}}{\partial Z}\circ s'\left(W'Y+b'\otimes 1_n^{\mathrm{T}}\right)\right]Y^{\mathrm{T}}+\left[\frac{\partial J_{AE}}{\partial Y}\circ s'\left(WX+b\otimes 1_n^{\mathrm{T}}\right)\right]X^{\mathrm{T}}\right\} \quad (14)$$

因为权重衰减项 J_{wd} 只作用于 W，所以 J_{wd} 关于 W 的偏导数可以表示为 $\frac{\partial J_{wd}}{\partial W}=W$。

对于稀疏约束项 J_{sparse}，关于 ρ_j^* 的偏导数可以表示为 $\frac{\partial J_{sparse}}{\partial \rho_j^*}=\frac{-\rho}{\rho_j^*}+\frac{1-\rho}{1-\rho_j^*}$。因为稀疏约束项是对隐含层进行稀疏，在使用反向传播方法时只需要对编码阶段反馈，所以 J_{sparse} 关于 b 和 W 的偏导为

$$\frac{\partial J_{sparse}}{\partial b}=\left[\frac{\partial J_{sparse}}{\partial \rho_j^*}\circ s'\left(WX+b\otimes 1_n^{\mathrm{T}}\right)\right]1_n \quad (15)$$

$$\frac{\partial J_{sparse}}{\partial W}=\left[\frac{\partial J_{sparse}}{\partial \rho_j^*}\circ s'\left(WX+b\otimes 1_n^{\mathrm{T}}\right)\right]X^{\mathrm{T}} \quad (16)$$

对于标签误差项 J_{label}，关于隐含层表示 Y 的偏导数可以表示为 $\frac{\partial J_{label}}{\partial Y}=Y\left(Y\hat{\beta}-T\right)$。相同地，标签误差项只作用于输入与隐含层之间，只产生它们之间的误差，所以 J_{label} 关于 b 和 W 的偏导为

$$\frac{\partial J_{label}}{\partial b}=\left[\frac{\partial J_{label}}{\partial Y}\circ s'\left(WX+b\otimes 1_n^{\mathrm{T}}\right)\right]1_n \quad (17)$$

$$\frac{\partial J_{label}}{\partial W}=\left[\frac{\partial J_{label}}{\partial Y}\circ s'\left(WX+b\otimes 1_n^{\mathrm{T}}\right)\right]X^{\mathrm{T}} \quad (18)$$

综合以上重构误差，权重衰减项，稀疏约束项和标签误差项的偏导数，可以得到 SLRAE 的目标函数 J_{LSRAE} 的偏导数为

$$\frac{\partial J_{LSRAE}}{\partial b}=\frac{\partial J_{AE}}{\partial b}+\lambda\frac{\partial J_{wd}}{\partial b}+\beta\frac{\partial J_{sparse}}{\partial b}+\nu\frac{\partial J_{label}}{\partial b} \quad (19)$$

$$\frac{\partial J_{LSRAE}}{\partial W}=\frac{\partial J_{AE}}{\partial W}+\lambda\frac{\partial J_{wd}}{\partial W}+\beta\frac{\partial J_{sparse}}{\partial W}+\nu\frac{\partial J_{label}}{\partial W} \quad (20)$$

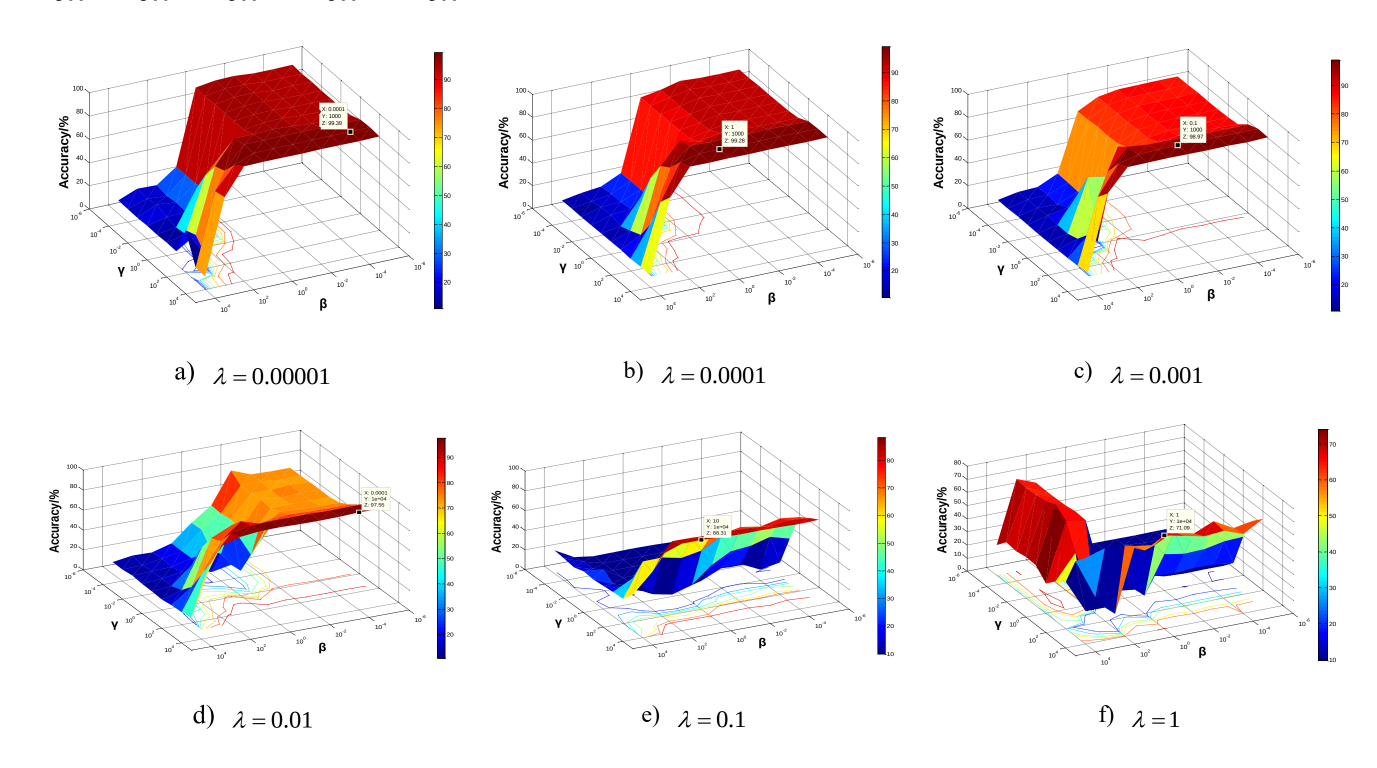
3 实验结果与分析

为了测试算法的性能，将提出的 SLRAE 与 AE，SAE，ELM 在一些数据集上做对比实验，然后应用到 softmax 分类器上进行分类。数据集主要包括 6 个 UCI 数据集，分别为 Pen Digits，Iris，Glass，Seeds，ISOLET，Page Blocks 和 1 个 USPS 数据集。具体数据信息如表 1 所示。算法是在 MATLAB R2014a 环境下进行，操作系统为 Windows 7，CPU 为 Intel(R) Core(TM) i3+3.40 GHz，4 GB 内存。

表 1 数据集的规格

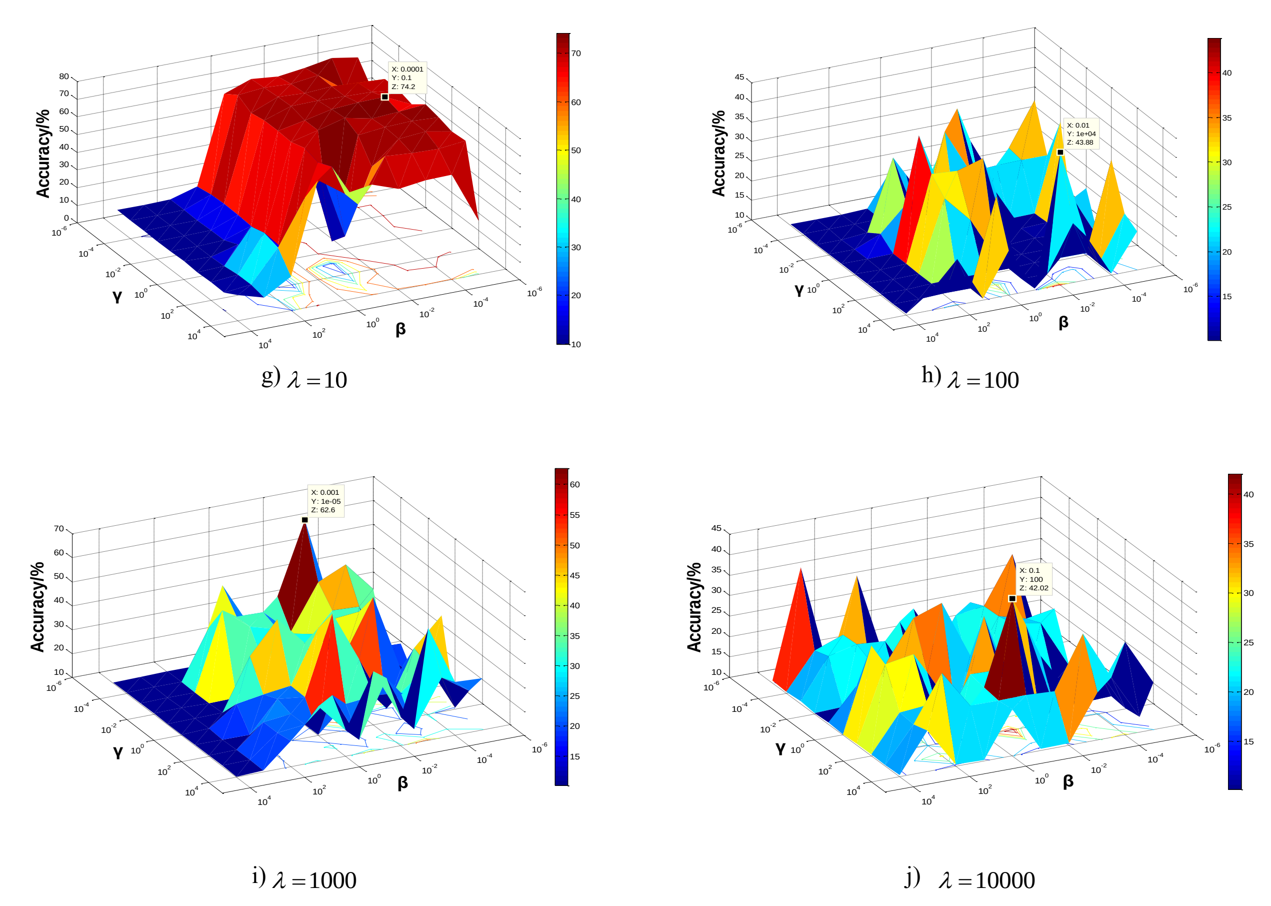
#	Data sets	Training	Testing	Total	Attribution	Classes
1	Pen Digits	7294	3498	10992	16	10
2	USPS	7291	2007	9298	256	10
3	Iris	75	75	150	4	3
4	Glass	214	140	354	10	7
5	Seeds	210	150	360	7	3
6	ISOLET	4000	1473	5473	10	54
7	Page Blocks	6238	1559	7797	617	26

SLRAE 算法的参数有 λ，β 和 γ，它们是用来控制公式中各项的相对重要性。对这些参数进行讨论，将这些参数以 $\{1\times10^{e}|e=-5,-4,-3,-2,-1,0,1,2,3,4\}$ 的规则变化。图 2 讨论了 SLRAE 使用 Pen Digits 数据库时的参数 $\lambda=0.00001$，$\lambda=0.0001$ 和 $\lambda=0.001$ 的性能。每张图表示一个不同的 $\lambda=0.01$ 的值，$\lambda=0.1$ 和 $\lambda=1$ 轴表示了参数的变化范围，彩条表示精度。可以从图 2 看出当 $\lambda=10$，$\lambda=100$，$\lambda=1000$ 时，SLRAE 达到最优的情况，分类的准确度达到 99.4%。$\lambda=10000$，β 和 γ 数量级相差大的原因是为了调整惩罚项达到同一个数量级。



a) $\lambda=0.00001$　b) $\lambda=0.0001$　c) $\lambda=0.001$

d) $\lambda=0.01$　e) $\lambda=0.1$　f) $\lambda=1$

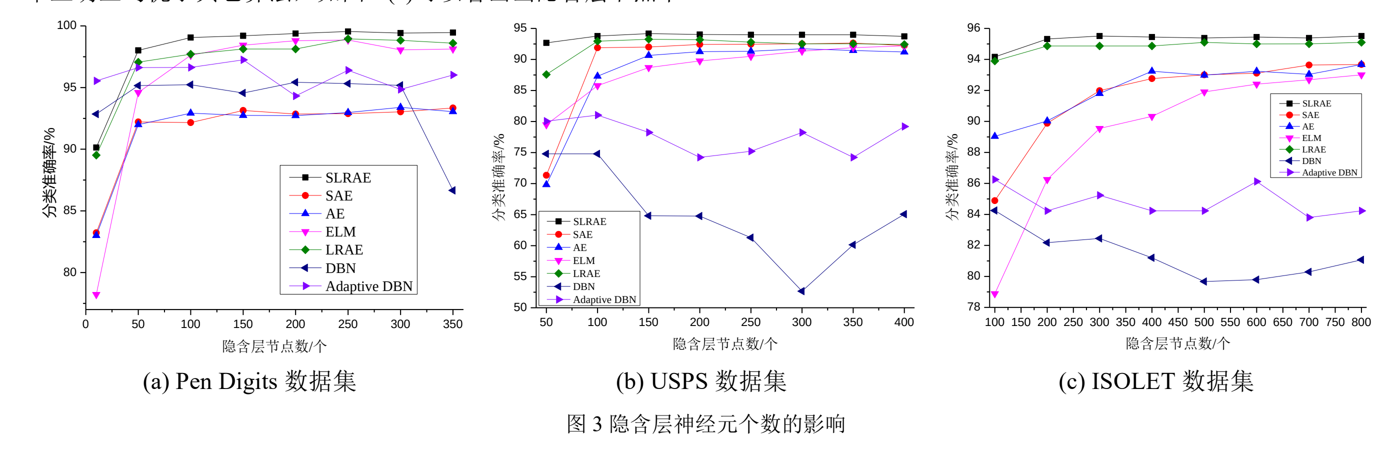


g) $\lambda=10$　　h) $\lambda=100$

i) $\lambda=1000$　　j) $\lambda=10000$

图 2　不同 λ，β 和 γ 下 SLRAE 的性能

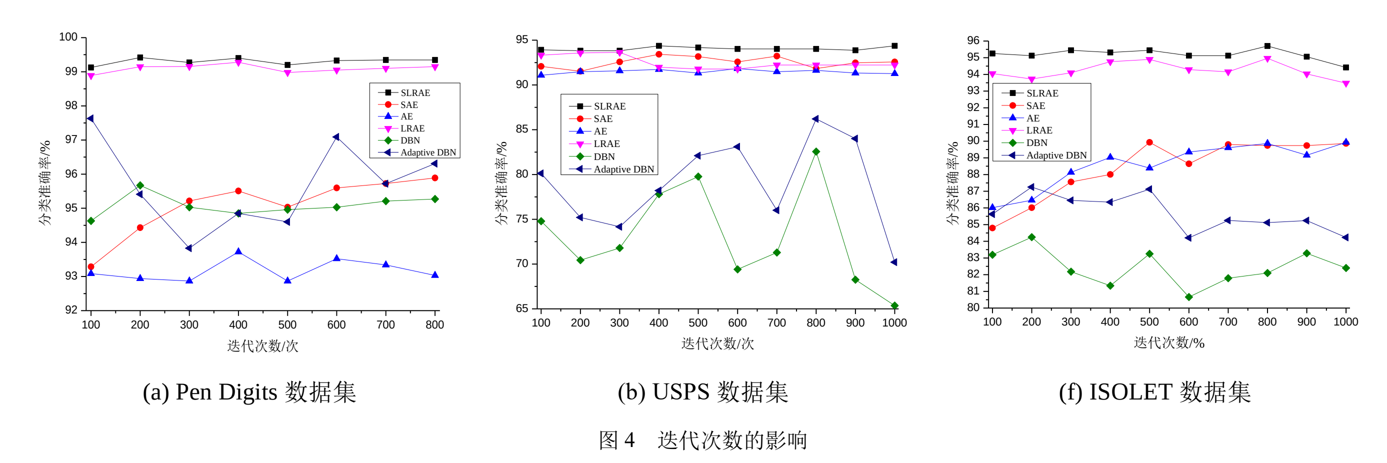
为了使比较公平，对比的网络结构有相同的网络结构(即有相同的深度和隐含层节点数)。以上实验得到分类结果最高的参数，之后我们研究神经元数目对性能的影响。本文首先为每个数据集设置最佳的模型参数，然后改变神经元隐含层节点数且固定其他参数。图 3 显示了 Pen Digits、USPS 和 ISOLET 数据集的分类结果。根据每个数据集的原始特征数来设置隐含层神经元的节点数。改变隐含层神经元的个数，SLRAE 在分类准确率上明显均优于其它算法，如图 3(a)可以看出当隐含层节点个数达到 100 时，SLRAE 用于分类的准确率就达到一个平稳的状态，分类的准确率接近于 100%。从图 3(b)中可以看出 SLRAE 用于分类的准确率差不多维持在 93%上下，实验结果一直比较平稳。SLRAE 用于分类的准确率是很明显高于其它算法。图 3(c)显示可以看出 SLRAE 用于分类的准确率是明显优于其它算法，不论隐含层节点个数为多少，SLRAE 的分类准确率差不多维持在 95.5%左右。



(a) Pen Digits 数据集　　(b) USPS 数据集　　(c) ISOLET 数据集

图 3 隐含层神经元个数的影响

在接下来的实验中研究不同的迭代次数对 SLRAE，SAE，AE，ELM，LRAE，DBN 以及 Adaptive DBN 的分类准确率的影响，将其进行对比实验。在刚刚隐含层节点个数对分类准确率的影响的实验中获得了每个方法在每个数据集上分类准确率最高的隐含层节点个数。图 4 是在 Pen Digits、USPS 和 ISOLET 数据集上迭代次数对各方法的影响。



(a) Pen Digits 数据集 (b) USPS 数据集 (f) ISOLET 数据集

图 4 迭代次数的影响

从图 4 可以看出在各迭代次数的情况下，SLRAE 用于分类的准确率都是很明显高于 SAE，AE，ELM，LRAE，DBN 以及 Adaptive DBN 这些算法的。从图 4(a)中看出迭代次数的对 SLRAE 的分类准确率的影响不大，分类准确率高大 99.41%。从图 4(b)中看出当迭代次数达到 400 时，SLRAE 用于分类的准确率达到稳定状态，分类准确率达到 94.4%。从图 4(c)中看出 SLRAE 用于分类的准确率是光滑的，并不随着迭代次数的增加而发生巨大改变，当迭代次数为 800 时，SLRAE 分类准确率最高为 95.7%。

表 2 各数据库上运行 100 次准确率范围

	AE	SAE	ELM	LRAE	DBN	Adaptive DBN	SLRAE
Pen Digits	92.73+0.63	94.58+0.73	94.30+1.33	99.18+0.25	93.07+2.03	96.67+1.64	**99.34+0.26**
	-0.66	-0.95	-1.34	-0.68	-2.06	-4.58	**-0.25**
USPS	85.87+2.87	87.79+1.95	89.75+0.78	90.55+0.89	70.35+0	73.49+1.49	**93.86+0.41**
	-2.81	-3.34	-1.11	-1.16	-0	-1.48	**-0.44**
Iris	96.82+1.85	96.91+1.76	93.68+4.99	96.83+1.84	97.33+0	95.83+1.5	**98.13+0.54**
	-0.82	-0.91	-5.68	-6.16	-0	-11.8	**-0.8**
Glass	83.87+14.73	96.93+3.07	90.06+8.07	85.63+13.44	98.07+0	84.46+14.14	**99.24+0.76**
	-19.85	-8.61	-15.26	-23.01	-0	-24.18	**-0.84**
Seeds	87.68+8.32	96.80+1.20	98.02+1.31	87.78+6.89	97.33+0	98.03+1.97	**98.67+1.33**
	-13.68	-1.47	-2.69	-11.78	-0	-5.36	**-1.34**
ISOLET	84.66+2.83	84.39+2.52	78.27+3.77	94.05+1.20	82.26+1.68	87.62+2.37	**94.25+1.06**
	-2.49	-2.99	-2.84	-1.31	-1.24	-2.54	**-1.18**
Page Blocks	88.35+5.46	86.33+1.31	92.41+2.49	92.71+1.71	86.33+0	88.55+4.85	**94.25+0.95**
	-18.35	-2.56	-3.16	-14.89	-0	-2.36	**-2.18**

表 2 中各数据的上一行表示的是平均准确率与到最大准确率的波动，下一行表示的是到最小准确率的波动。通过表 2 可以看出 SLRAE 方法在各数据集上的分类准确率的波动都较小，有时波动大于 DBN，DBN 在数据集 USPS、Iris、Glass、Seeds 和 Page Blocks 上的波动为 0，但是 SLRAE 的平均分类准确率是高于 DBN 以及其它算法的，所以 SLRAE 用于分类的优势是明显高于其他算法的。

4 结束语

在本文中，将稀疏和标签正则化加入到自动编码机中提出一个半监督自动编码机(SLRAE)。添加稀疏约束项可以有效地提取隐含层结构，学习更复杂的非线性函数，更准确地提取样本的本质特征，添加标签约束项通过减小实际标签和期望标签的误差，使分类精度高于有监督和无监督模型。单独使用稀疏约束项没有用到目标信息不利于用到分类中，单独的加入标签项导致学习到的特征不能很好的表示原始的数据，泛化能力变弱，模型容易过度拟合。所以针对这一问题，为了发挥无监督学习与监督学习的优势，提出了加入稀疏约束项和标签项的半监督自动编码机，通过对自编码参数和分类参数的同时优化来达到既可以保留原数据特征，保证模型的泛化能力，又可以实现较好分类效果的目的。

参考文献：

[1] Hinton G E, Osindero S, Teh Y W. A fast learning algorithm for deep belief nets [J]. Neural Computation, 2014, 18 (7): 1527-1554.

[2] Bengio Y. Learning deep architectures for AI [M]. [S. l.] : Now Publishers, 2009.

[3] Li Deng, Dong Yu. Deep learning: methods and applications [J].

Foundations & Trends in Signal Processing, 2014, 7 (3): 197-387.

[4] Bengio Y, Lamblin P, Popovici D, *et al.* Greedy layer-wise training of deep networks [C]// Advances in Neural Information Processing Systems. 2007: 153-160.

[5] Sun Wenjun, Shao Siyu, Zhao Rui, *et al.* A sparse auto-encoder-based deep neural network approach for induction motor faults classification [J]. Measurement, 2016, 89: 171-178.

[6] Vincent P, Larochelle H, Bengio Y, *et al.* Extracting and composing robust features with denoising autoencoders [C]// Proc of International Conference on Machine Learning. New York: ACM Press, 2008: 1096-1103.

[7] Vincent P, Larochelle H, Lajoie I, *et al.* Stacked denoising autoencoders: learning useful representations in a deep network with a local denoising criterion [J]. Journal of Machine Learning Research, 2010, 11 (12): 3371-3408.

[8] Rifai S, Mesnil G, Vincent P, *et al.* Higher order contractive auto-encoder [M]// Machine Learning and Knowledge Discovery in Databases. Berlin: Springer, 2011: 645-660.

[9] Liu Yanan, Feng Xiaoqing, Zhou Zhiguang. Multimodal video classification with stacked contractive autoencoders [J]. Signal Processing, 2016, 120: 761-766.

[10] Du Bo, Wei Xiong, Jia Wu, *et al.* Stacked Convolutional Denoising Auto-Encoders for Feature Representation [J]. IEEE Trans on Cybernetics, 2016, 99: 1-11.

[11] Liu Weifeng, Ma Tengzhou, Xie Qiangsheng, *et al.* LMAE: a large margin auto-encoders for classification [J]. Signal Processing, 2017, 141: 137-143.

[12] http://www. datatang. com/data/11927 [EB/OL].

[13] Blake C, Merz C. Uci repository of machine learning databases [DB//OL]. 1998.

[14] Liu Ping, Han Shizhong, Meng Zibo, *et al.* Facial expression recognition via a boosted deep belief network [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. Washington DC: IEEE Computer Society, 2014: 1805-1812.

[15] Kamada S, Ichimura T. An adaptive learning method of deep belief network by layer generation algorithm [C]// Proc of Region 10 Conference. 2017: 2967-2970.

[16] Zou Qiang, Fang Hui, Liu Fei, *et al.* Comparative study of distance discriminant analysis and BP neural network for identification of rapeseed cultivars using visible//near infrared spectra [C]// IFIP Advances in Information & Communication Technology, 2017, 347: 124-133.

[17] Huang GuangBin, Zhu QinYu, Siew C K. Extreme learning machine: Theory and applications [J]. Neurocomputing, 2006, 70 (1–3): 489-501.

[18] Huang Guangbin, Wang Dianhui, Lan Yuan. Extreme learning machines: a survey [J]. International Journal of Machine Learning & Cybernetics, 2011, 2 (2): 107-122.

[19] 林雨. 极限学习机与自动编码器的融合算法研究 [D]. 长春: 吉林大学, 2016. (Lin Yu. Research on fusion algorithm of extreme learning machine and auto-encoder [D]. Changchun: Jilin University, 2016.)